

# Enhanced thermionic-dominated photoresponse in graphene Schottky junctions

Joaquin F. Rodriguez-Nieva<sup>1</sup>, Mildred S. Dresselhaus<sup>1,2</sup>, and Justin C. W. Song<sup>3</sup>

<sup>1</sup> *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA*

<sup>2</sup> *Department of Electrical Engineering and Computer Science,*

*Massachusetts Institute of Technology, Cambridge, MA 02139, USA and*

<sup>3</sup> *Walter Burke Institute for Theoretical Physics and Institute of Quantum Information and Matter, California Institute of Technology, Pasadena, CA 91125 USA*

Vertical heterostructures of van der Waals materials enable new pathways to tune charge and energy transport characteristics in nanoscale systems. We propose that graphene Schottky junctions can host a special kind of photoresponse which is characterized by strongly coupled heat and charge flows that run vertically out of the graphene plane. This regime can be accessed when vertical energy transport mediated by thermionic emission of hot carriers overwhelms electron-lattice cooling as well as lateral diffusive energy transport. As such, the power pumped into the system is efficiently extracted across the entire graphene active area via thermionic emission of hot carriers into a semiconductor material. Experimental signatures of this regime include a large and tunable internal responsivity  $\mathcal{R}$  with a non-monotonic temperature dependence. In particular,  $\mathcal{R}$  peaks at electronic temperatures on the order of the Schottky potential  $\phi$  and has a large upper limit  $\mathcal{R} \leq e/\phi$  ( $e/\phi = 10$  A/W when  $\phi = 100$  meV). Our proposal opens up new approaches for engineering the photoresponse in optically-active graphene heterostructures.

Keywords: graphene; hot carriers; Schottky junction; photocurrent

Vertical heterostructures comprising layers of van der Waals (vdW) materials have recently emerged as a platform for designer electronic interfaces [1]. Of special interest are heterostructures which feature tunable interlayer transport characteristics, as exemplified by g/X Schottky junctions [2–10]; here ‘g’ denotes graphene, and X is a semiconductor material, such as Si, MoS<sub>2</sub> or WSe<sub>2</sub>. These junctions are characterized by Schottky barriers  $\phi$  that span two orders of magnitude  $\phi \approx 0.01 - 1$  eV and exhibit in situ control through applied bias or by using gate potentials [7–11]. The wide range of  $\phi$  achievable across the g/X interface, combined with the unique graphene photoresponse mediated by long-lived hot carriers (elevated electronic temperatures,  $T_g$ , different from those of the lattice,  $T_0$  [12–17]), make graphene Schottky junctions a prime target for accessing novel vertical energy transport regimes [18].

Here we show that specially designed graphene Schottky junctions can host an enhanced thermionic-dominated photoresponse driven by strongly coupled charge and energy currents. Such photoresponse proceeds, as illustrated in Fig. 1a, via the thermionic emission of graphene hot carriers with energy larger than the Schottky barrier. At steady state, an equal number of cold carriers are injected at the Fermi surface through an ohmic contact, giving a net flow of heat  $\mathcal{J}_q^\perp$  out of the graphene electronic system balancing the energy pumped into the system.

Strikingly, thermionic emission yields strong heat transport running vertically out of the hot electron system, which dominates over more conventional electronic cooling channels, e.g. electron-lattice cooling. Indeed, we find that  $\mathcal{J}_q^\perp$  can be significant in graphene (see Fig. 1c) when  $k_B T_g \approx \phi/2$ , dominating over acoustic and opti-

cal phonon cooling [12, 13] in pristine graphene Schottky junctions;  $\mathcal{J}_q^\perp$  also overwhelms in-plane (lateral) diffusive energy transport. We find that the values of  $\mathcal{J}_q^\perp$  are competitive with disorder-assisted cooling [14–16] in more dirty devices.

Graphene is essential to our proposal due to a unique combination of electronic characteristics. First, fast intraband Auger-type scattering [19, 20] allows the absorbed photon energy flux,  $\mathcal{J}_q^{\text{in}}$ , to be efficiently captured as heat by ambient carriers in graphene; this process results in a thermalized hot carrier distribution [19, 20]. Second, graphene is characterized by slow electron-lattice cooling mechanisms [12–17] which enables  $T_g > T_0$  to drive a strong thermionic current. This is due to the large optical phonon energy in graphene [12, 13] as well as the weak electron-acoustic phonon coupling [for a detailed comparison between cooling rates, see Eq.(8) below]. Third, the bias and gate-tunable work function allows an experimentally accessible way to optimize device operation, predicted to occur at  $k_B T_g \approx \phi/2$ , for a range of technologically achievable temperatures and barrier materials. Indeed, whereas other Schottky junctions (e.g. Au/Si, Ag/Si) may also display vertical energy currents, their large Schottky barriers ( $\phi \approx 1$  eV) and fast electron-lattice cooling render the thermionic-dominated regime impractical for these systems.

An important optoelectronic figure of merit is the conversion between incoming photon energy flux,  $\mathcal{J}_q^{\text{in}}$ , and detected photocurrent,  $\mathcal{J}_e^\perp$ , encoded in the (internal) responsivity  $\mathcal{R} = \mathcal{J}_e^\perp / \mathcal{J}_q^{\text{in}}$ . Importantly, our model yields a large upper limit for  $\mathcal{R}$ . Indeed, energy  $\phi$  is transported per carrier extracted across the g/X junction (Fig. 1a) yielding a limiting internal responsivity  $\mathcal{R}_0$ , occurring in the thermionic-dominated regime (i.e.,  $\mathcal{J}_q^\perp = \mathcal{J}_q^{\text{in}}$ ) given

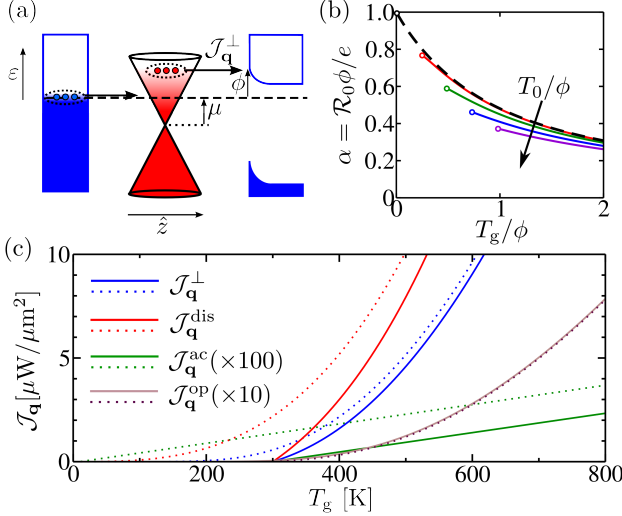


FIG. 1. (a) Schematic of thermionic-dominated heat transport in a graphene Schottky junction: hot carriers with energies  $\epsilon$  close to the Schottky barrier height  $\phi$  are thermionically emitted into a semiconductor material in the out-of-plane direction  $\hat{z}$ , while cold carriers are injected through an Ohmic contact at the Fermi level,  $\mu$ , generating a net vertical heat current  $\mathcal{J}_q^\perp$ . (b) Normalized ideal responsivity,  $\alpha = \mathcal{R}_0\phi/e$ , is shown as a function of normalized graphene electronic temperature,  $k_B T_g/\phi$ , with  $e$  the electron charge [see Eq.(1)]. Curves are obtained for normalized ambient temperature  $k_B T_0/\phi = 0, 0.25, 0.5, 0.75, 1.0$ , indicated with different colors for increasing  $T_0$ ; characteristic  $\mathcal{R}_0 \approx e/\phi \sim 10$  A/W can be large (for  $\phi = 100$  meV, see text). (c) Thermionic cooling channel,  $\mathcal{J}_q^\perp$ , compared with acoustic phonon cooling (clean case),  $\mathcal{J}_q^{\text{ac}}$ , disorder-assisted cooling,  $\mathcal{J}_q^{\text{dis}}$ , and optical phonon cooling  $\mathcal{J}_q^{\text{op}}$  [see text and Eq.(8)], shown for ambient temperature  $T_0 = 300$  K (solid lines) and  $T_0 = 0$  (dotted lines). Note that  $\mathcal{J}_q^\perp$  overwhelms  $\mathcal{J}_q^{\text{ac}}$  and  $\mathcal{J}_q^{\text{op}}$  (clean case) and is competitive with  $\mathcal{J}_q^{\text{dis}}$  (dirty case). Parameter values used:  $\mu = \phi = 100$  meV,  $k_F \ell = 50$ , and  $G_0 = 10$  mS/ $\mu\text{m}^2$ , see Eqs. (4) and (8).

by

$$\mathcal{R} \leq \mathcal{R}_0, \quad \mathcal{R}_0 = \frac{e}{\phi} \times \alpha(\tilde{T}_g, \tilde{T}_0). \quad (1)$$

Here  $e$  is the electron charge,  $\tilde{T}_{g,0} = k_B T_{g,0}/\phi$  are the dimensionless graphene and ambient temperatures (temperature of the lattice and semiconductor), respectively, and  $\alpha$  is a dimensionless function (see text below) plotted in Fig. 1b. The function  $\alpha$  can take values close to unity, allowing  $\mathcal{R}_0$  to be on the order of  $e/\phi = 10$  A/W, for  $\phi = 100$  meV. For a discussion of net values of  $\mathcal{R}$  in Eq.(1), see Fig. 2c.

Since the incident photon energy  $\hbar\omega$  (e.g. in the visible) can be many multiples of  $\phi$ , we anticipate that g/X Schottky photodetectors can provide significant gains in the internal responsivity compared to those in conventional (photovoltaic-based) photodetectors, which are limited by  $\mathcal{R}^{\text{PV}} \leq e/\hbar\omega$  [21]. In particular, the ultra-fast

electron energy relaxation times in graphene yield *multiple* hot carriers per absorbed photon [19, 20], in stark contrast to photovoltaic-based schemes that yield a single electron-hole pair per absorbed photon. Naturally, the external responsivity of the device is also affected by the absorption coefficient of the photoactive material. Whereas the absorption coefficient of 2.3% per layer in graphene [22] is small compared, for instance, to typical values of 10-50% in Si [23], this small value can be increased using optical waveguides [3] and plasmon enhanced absorption [24]. These external enhancement mechanisms will not be discussed here.

In addition, g/X photodetectors also enable a boosted photoresponse compared to previous photothermoelectric-based schemes [17]. Indeed, the vertical structure allows to circumvent lateral electronic heat diffusion, which drastically reduces the operating electronic temperatures and efficiencies in photothermoelectric-based schemes.

Another important feature of the g/X photoresponse is the possibility of using the temperature dependence of  $\mathcal{R}$  as a diagnostic of thermionic-dominated photoresponse. Including losses to the lattice via disorder-assisted cooling, we find that  $\mathcal{R}$  is non-monotonic, peaking at an optimal operating hot carrier temperature  $k_B T_g \approx \phi/2$  (Fig. 2c). Since  $T_g$  can be controlled by the incident light power and  $\phi$  via gate voltage, non-monotonic  $\mathcal{R}$  as a function of  $T_g$  provides an easily accessible experimental signature of the strongly coupled charge and energy thermionic transport that is engineered across the g/X interface. Indeed, non-monotonic temperature behavior does not occur in the photovoltaic-based devices, where responsivity is mainly independent of pump power or photon intensity.

We begin by modeling vertical transport across the g/X device, as depicted in Fig. 1a. To describe thermionic transport over the barrier, we adopt a quasielastic but momentum non-conserving approximation [25]. This approximation is valid because, at high  $T$ , a number of momentum scattering mechanisms at the Schottky junction are possible, such as scattering by defects, intrinsic phonons and substrate phonons. Furthermore, the typical energy exchange in these processes is small on the Schottky barrier scale. As a result, we generically write the electron and heat current across the g/X Schottky junction as

$$\begin{aligned} \begin{bmatrix} \mathcal{J}_e^\perp \\ \mathcal{J}_q^\perp \end{bmatrix} &= \int_{-\infty}^{\infty} d\epsilon \begin{bmatrix} e \\ \epsilon \end{bmatrix} G(\epsilon) \left[ f\left(\frac{\epsilon}{k_B T_g}\right) - f\left(\frac{\epsilon}{k_B T_0}\right) \right], \\ G(\epsilon) &= \frac{2\pi e^2}{\hbar} D_g(\epsilon) D_c(\epsilon) |T(\epsilon)|^2. \end{aligned} \quad (2)$$

Here  $G(\epsilon)$  is a parameter with units of electrical conductance which characterizes the Schottky interface (see discussion below),  $D_g$  ( $D_c$ ) is the density of states of graphene (the conduction band of the semiconductor),  $f(x) = 1/(e^x + 1)$  is the Fermi distribution function,

$T(\varepsilon)$  is the energy-dependent tunneling transition matrix element between graphene and semiconductor electronic states, and energies  $\varepsilon$  are referenced from the Fermi energy  $\mu$  (see Fig. 1a). The function  $T(\varepsilon)$  contains all the microscopic information about the relevant mechanisms that couple graphene with material X, such as phonons or hot-spots formed by defects.

Two important assumptions are present in Eq.(2). First, we neglected hole transport between graphene and the valence band of X assuming that the barrier height for hole transport is much larger than the corresponding one for electron transport. Secondly, we assume that the Fermi level and the temperature in graphene and X are spatially fixed. In a more realistic scenario, the pumping power may cause the temperature and Fermi level to spatially vary in the out-of plane direction. In this case, both quantities need to be determined self-consistently by appropriate balance equations. However, these do not introduce any new qualitative features to our simplified model.

When light heats graphene electrons so that  $T_g > T_0$ , Eq.(2) describes the short-circuit charge current (photocurrent) and the energy current flow, shown schematically in Fig. 1c. At steady state,  $T_g$  is determined by energy balance of the incident absorbed power in graphene,  $\mathcal{J}_q^{\text{in}}$ , and the energy being dissipated by the graphene electronic system,  $\mathcal{J}_q^{\text{out}}$ , that includes both the thermionic channel,  $\mathcal{J}_q^\perp$ , and other dissipative channels,  $\mathcal{J}_q^{\text{loss}}$  (e.g. electron-lattice cooling, and diffusive heat transport discussed below). Explicitly, we have

$$\mathcal{J}_q^{\text{in}} = \mathcal{J}_q^{\text{out}} = \mathcal{J}_q^\perp(T_g, T_0) + \mathcal{J}_q^{\text{loss}}(T_g, T_0), \quad (3)$$

where we have fixed  $T_0$  to the temperature of the ambient environment, i.e. there is no backflow of hot electrons into graphene. The latter assumption results from the large heat capacity and fast electron-lattice cooling in highly doped semiconductors such as Si [26]. In what follows, we shall analyze the energy/charge characteristics of g/X Schottky junctions as a function of  $T_g$  and  $T_0$ ; naturally the  $T_g$  values displayed can be attained via a suitably chosen  $\mathcal{J}_q^{\text{in}}$ .

The depletion width, for example in g/Si interfaces[6], can be many times larger than the electron wavelength. As a result, only electrons with energies above the effective barrier  $\phi$  formed at the g/X interface contribute to the current; in this way, the photocurrent is thermally activated. Here we adopt a phenomenological approach to capture the essential physics independent of the microscopic details of the device. To this end, we approximate  $G(\varepsilon) = G_0\Theta(\varepsilon - \phi)$  in Eq.(2), with  $\Theta$  the step-function, in order to aggregate the microscopics of the junction into a single variable that can be easily measured in experiments. This approach does not describe field emission, which is considered negligible because we are limiting our discussion to the zero bias behavior, i.e.

the closed circuit photocurrent. We emphasize that this approximation does not affect the qualitative behavior of  $\mathcal{J}_e^\perp$ ,  $\mathcal{J}_q^\perp$  or  $\alpha$  for the range of temperatures of interest,  $T_g \lesssim \phi$ ; further, this approximation represents a conservative estimate of the particle current, since  $G(\varepsilon)$  is typically a monotonically increasing function due to the larger density of states available for scattering at larger  $\varepsilon$  in graphene. As a side remark, we note that  $G_0$  is *not* the zero-bias junction conductance; this quantity is suppressed by a factor  $e^{-\phi/T}$  with respect to  $G_0$ , as discussed in the paragraph following Eq.(10).

Using a step-like transmission, heat and charge currents can then be expressed in terms of non-dimensional integrals by defining  $x = \varepsilon/\phi$  in Eq.(2), yielding

$$\mathcal{J}_e^\perp = \frac{G_0\phi}{e} \int_1^\infty dx \Delta f(x), \quad \mathcal{J}_q^\perp = \frac{G_0\phi^2}{e^2} \int_1^\infty dx x \Delta f(x), \quad (4)$$

where  $\Delta f(x) = f(x/\tilde{T}_g) - f(x/\tilde{T}_0)$ . It is straight-forward to show that the integrals on the right-hand side of Eq.(4) are related to the well-known complete Fermi integrals,

$$F_k(\xi) = \int_0^\infty dx \frac{x^k}{e^{x-\xi} + 1}, \quad (5)$$

via  $\int_1^\infty dx x^n f(x/\tilde{T}) = \sum_{k=0}^n \binom{n}{k} \tilde{T}^{k+1} F_k(-1/\tilde{T})$ . In

the low temperature regime,  $\tilde{T} \ll 1$ , the value of  $F_k$  behaves as  $F_k(-1/\tilde{T}) \approx \Gamma(k+1)e^{-1/\tilde{T}}$ , with  $\Gamma$  the Gamma function. In the high temperature regime,  $\tilde{T} \gg 1$ ,  $F_k$  takes values  $F_0(-1/\tilde{T}) \approx \ln(2)$  and  $F_1(-1/\tilde{T}) \approx \pi^2/12$ .

A key feature of thermionic-dominated [ $\mathcal{J}_q^{\text{loss}} = 0$ ] transport is the strong charge-energy current coupling manifested in  $\mathcal{R}_0$ . Using Eq.(4) above, we obtain Eq.(1) with  $\alpha$  given by

$$\alpha(\tilde{T}_g, \tilde{T}_0) = \frac{\int_1^\infty dx \Delta f(x)}{\int_1^\infty dx x \Delta f(x)}. \quad (6)$$

The general behavior of  $\alpha(\tilde{T}_g, \tilde{T}_0)$  can be most easily understood by first setting  $T_0 = 0$ . In this case,  $\alpha$  in Eq.(1) adopts the simple form

$$\alpha_0(\tilde{T}_g) \equiv \alpha(\tilde{T}_g, \tilde{T}_0 = 0) = \left[ 1 + \tilde{T}_g \frac{F_1(-1/\tilde{T}_g)}{F_0(-1/\tilde{T}_g)} \right]^{-1}. \quad (7)$$

Importantly,  $\alpha_0$  is a decreasing function of temperature  $\tilde{T}_g$ , as shown in the black dashed curve of Fig. 1b. In particular, for  $\tilde{T}_g \ll 1$ ,  $\alpha_0$  takes values  $\alpha_0 \sim (1 + \tilde{T}_g)^{-1}$  of order unity, and for  $\tilde{T}_g \gg 1$ ,  $\alpha_0$  decreases with inverse temperature as  $\alpha_0 \approx 12\log(2)/(\pi^2\tilde{T}_g)$  (see Fig. 1b). This latter fact means that, although  $\mathcal{R}_0$  is expressed in units of  $e/\phi$  in Eq.(1),  $\mathcal{R}_0$  cannot grow indefinitely by making  $\phi$  smaller;  $\mathcal{R}_0$  reaches a saturating value  $\mathcal{R}_0 \approx 12\log(2)e/(\pi^2k_B T_g)$  for  $T_g \gg \phi$ , as shown in Fig. 1c.

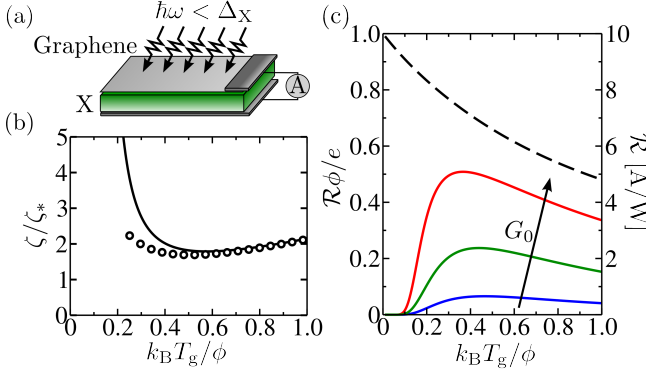


FIG. 2. (a) Schematic of a hot carrier photodetector formed by a graphene-X Schottky junction, where X is a semiconductor material with bandgap  $\Delta_X$ . Here we consider graphene as the photoactive material for absorption, i.e. photon energy  $\hbar\omega < \Delta_X$ . (b) Ratio  $\zeta = \mathcal{J}_q^{\text{loss}}/\mathcal{J}_q^\perp$  modeled via Eq.(10) as a function of  $T_g$ , with  $T_0 = 0$  (solid line) and  $T_0 = 300$  K (empty circles). Note that the latter has a smaller range,  $T_g \geq T_0$ . (c) The responsivity  $\mathcal{R}$  (solid lines) for the g/X junction exhibits a non-monotonic electronic temperature dependence peaking at  $T_g \approx \phi/2k_B$ , shown for  $\zeta$  modeled in Eq.(10) [panel (b)] with  $\zeta_* = 5, 1, 0.2$  (blue, green, red, respectively); dimensionless  $\mathcal{R}$  is shown on the left vertical axis. The dashed line indicates an ideal case  $\mathcal{R} = \mathcal{R}_0$ . Here we used values  $G_0 = 2, 10, 50$  mS/ $\mu\text{m}^2$ ,  $\mu = \phi = 100$  meV and  $k_F\ell = 50$  yielding sizable  $\mathcal{R}$  (right vertical axis).

For finite values of  $T_0$ , the qualitative behavior of  $\alpha$  does not depart significantly from that of  $\alpha_0$ . As shown in Fig. 1b, where  $\alpha(\tilde{T}_g, \tilde{T}_0)$  is obtained by numerical integration of Eq.(6) for different values of  $\tilde{T}_0$ , the function  $\alpha$  largely follows the  $\alpha_0$  curve and only shifts slightly from  $\alpha_0$  with increasing  $\tilde{T}_0$ . Further, the range of operating hot electron temperatures is now smaller,  $\tilde{T}_g \geq \tilde{T}_0$ , as illustrated in Fig. 1b by curves that now start at  $\tilde{T}_g = \tilde{T}_0$ . Although  $\alpha$  is finite at  $\tilde{T}_g = \tilde{T}_0$ , there is no net current at equal temperatures (as indicated by the empty circles at the beginning of the curves in Fig. 1b); a non-vanishing  $\alpha$  at  $\tilde{T}_g = \tilde{T}_0$  arises from the differential ratio that characterizes the responsivity  $\mathcal{R}$ .

Considering losses, Eq.(4) yields  $\mathcal{J}_q^\perp$  that can be sizable (see blue curves in Fig. 1c). In plotting Fig. 1c, we numerically integrated Eq.(4) and used  $\phi = 100$  meV and  $G_0 = 10$  mS/ $\mu\text{m}^2$  (see below for  $G_0$  estimates). Further, we find that  $\mathcal{J}_q^\perp$  compares favorably with intrinsic electron-lattice cooling in graphene: (i) single-acoustic phonon cooling for pristine graphene  $\mathcal{J}_q^{\text{ac}}$  (green curves), (ii) optical phonon cooling  $\mathcal{J}_q^{\text{op}}$  (magenta curves), and (iii) disorder-assisted cooling  $\mathcal{J}_q^{\text{dis}}$  (red curves), where we consider the degenerate limit ( $\mu \gg T_g$ ) for all cases [12–

14]:

$$\begin{aligned} \mathcal{J}_q^{\text{ac}} &= \gamma_{\text{ac}}(T_g - T_0), & \mathcal{J}_q^{\text{dis}} &= \gamma_{\text{dis}}(T_g^3 - T_0^3), \\ \mathcal{J}_q^{\text{op}} &= \gamma_{\text{op}}[N(\hbar\omega_{\text{op}}/k_B T_g) - N(\hbar\omega_{\text{op}}/k_B T_0)]. \end{aligned} \quad (8)$$

Here the prefactors are  $\gamma_{\text{ac}} = \hbar D^2 \mu^4 k_B / 8\pi \rho (\hbar v_F)^6$ ,  $\gamma_{\text{dis}} = 2D^2 \mu^2 k_B^3 / \rho c^2 \hbar (\hbar v_F)^4 k_F \ell$  and  $\gamma_{\text{op}} = \hbar^2 \omega_{\text{op}}^3 / 24\pi \rho a^4 v_F^2$ , with  $D$  the deformation potential,  $\rho$  the graphene mass density,  $k_F \ell$  the dimensionless disorder parameter,  $a$  the lattice constant,  $\omega_{\text{op}}$  the optical phonon frequency, and  $N(x)$  the Bose distribution. For the cooling mechanisms, we used  $\mu = 100$  meV,  $D = 20$  eV,  $\rho = 7.6 \cdot 10^{-7}$  kg/m<sup>2</sup>,  $k_F \ell = 50$ ,  $a = 1.4$  Å and  $\hbar\omega_{\text{op}} = 0.2$  eV. Indeed,  $\mathcal{J}_q^\perp$  overwhelms both  $\mathcal{J}_q^{\text{ac}}$  and  $\mathcal{J}_q^{\text{op}}$ , and is competitive with  $\mathcal{J}_q^{\text{dis}}$ , as shown in Fig. 1c.

The hot carrier thermionic cooling channel,  $\mathcal{J}_q^\perp$ , and the strong charge-energy current coupling it produces [Eq.(1)], can manifest itself in large and non-monotonic responsivities in g/X photodetectors (Fig. 2a). Accounting for energy balance in Eq.(3) we find a net responsivity given by

$$\mathcal{R} = \frac{\mathcal{R}_0}{1 + \zeta}, \quad \zeta = \frac{\mathcal{J}_q^{\text{loss}}}{\mathcal{J}_q^\perp}, \quad (9)$$

where  $\zeta$  quantifies losses. To estimate  $\zeta$  for actual devices, we consider the disorder-assisted cooling power in graphene [14],  $\mathcal{J}_q^{\text{loss}} \approx \mathcal{J}_q^{\text{dis}} = \gamma_{\text{dis}}(T_g^3 - T_0^3)$  as an illustrative example, see Eq.(8). Adopting the same procedure as described above, we find

$$\zeta(\tilde{T}_g, \tilde{T}_0) = \zeta_* \times \frac{\tilde{T}_g^3 - \tilde{T}_0^3}{\int_1^\infty dx x \Delta f}, \quad \zeta_* = \frac{e^2 \gamma_{\text{dis}} \phi}{G_0}, \quad (10)$$

where the characteristic  $\zeta$  is set by  $e^2 \gamma_{\text{dis}} \phi / G_0$ . As expected, increasing the prefactor  $\gamma_{\text{dis}}$ , for  $\mathcal{J}_q^{\text{dis}}$ , increases the losses to phonon scattering embodied in  $\zeta$ . Alternatively, increasing the conductance across the g/X interface enhances the thermionic channel.

In calculating  $\mathcal{R}$  in Eq.(9), we use the same parameter values as in Fig. 1:  $\mu = \phi = 100$  meV and  $k_F \ell = 50$ . The value of  $G_0$  can be estimated from conductance measured in the dark state,  $G_D$ , obtained in actual g/X devices at equilibrium  $T_g = T_0$  (for example g/Si Schottky junctions in Refs. [7, 8]). Indeed, under an infinitesimally small potential bias  $\delta V_b$ , we can approximate  $\Delta f$  in Eq.(6) as  $\Delta f(x) = [e^x / (e^x + 1)^2] \times e \delta V_b / T_0$  due to the small chemical potential difference  $e \delta V_b$  between G and X. Integrating over  $x$  in Eq.(6), we obtain  $G_D = G_0 / (1 + e^{\phi/T_0})$ . In a typical scenario  $T_0 \ll \phi$ , the conductance in the dark state is exponentially suppressed with increasing temperature as  $G_D \approx G_0 \exp(-\phi/T_0)$ , in agreement with the qualitative behavior observed in Refs. [2–4, 7–10]. To give an estimate of the range of conductances achievable in g/X devices,  $G_D$  in these experiments report  $G_D \sim$

0.1 – several  $\times \mu\text{S}/\mu\text{m}^2$  for  $\phi_{\text{Si}} \sim 0.3\text{ eV}$  with  $T_0$  at room temperature. This gives  $G_0$  in the 1 – 100  $\text{mS}/\mu\text{m}^2$  ballpark [for  $\mathcal{R}$  in Fig. 2c, we used  $G_0 = 2, 10, 50\text{ mS}/\mu\text{m}^2$ , which correspond to  $\zeta_* = 5, 1, 0.2$ , see Eq.(10)].

As shown in Fig. 2b,  $\zeta$  exhibits a clear non-monotonic dependence on  $T_g$  characterized by two regimes: (i) small  $\tilde{T}_g \ll 1$ ,  $\mathcal{J}_q^\perp$  is exponentially suppressed by the transport barrier  $\phi$ , thus  $\mathcal{J}_q^{\text{dis}}$  dominates, (ii) large  $\tilde{T}_g \gg 1$ , we find that  $\mathcal{J}_q^\perp$  scales as  $T_g^2$ , and rises less steeply than the  $T_g^3$  power law of supercollision cooling. Hence, there is a “sweet spot” for observing a competitive thermionic channel  $\mathcal{J}_q^\perp$ . The optimal value occurs for temperatures  $T_g/\phi \approx 0.5$  (see Fig. 2b), with minimum  $\zeta_{\text{min}} \approx 1.85 \times \zeta_*$ . This can be estimated from Eq.(10) in the limit  $\tilde{T}_g \ll 1$  and  $T_0 = 0$ , where the above-mentioned optimal values are obtained from minimization of the equation  $\zeta/\zeta_* \approx \tilde{T}_g^2 e^{1/\tilde{T}_g}$  (Fig. 2b).

The responsivity  $\mathcal{R}$  in Eq.(9) mirrors  $\zeta$  to display a non-monotonic dependence on  $T_g$ , peaking at a temperature  $k_B T_g \approx \phi/2$ , as shown in Fig. 2c. Peak responsivities in the range  $\mathcal{R} \approx 1 - 10\text{ A/W}$  are obtained within our model. Indeed, for large  $G_0 = 50\text{ mS}/\mu\text{m}^2$  (corresponding to  $\zeta_* = 0.2$ ),  $\mathcal{R}$  starts to approach the ideal case,  $\mathcal{R} = \mathcal{R}_0$  (dashed black line). The non-monotonic dependence of  $\mathcal{R}$  as a function of  $T_g$  provides a clear fingerprint of the competition between thermionic energy transport and conventional electron-phonon cooling. Since the Schottky barrier heights can be tuned by the applied gate voltage, the peak temperature  $k_B T_g \approx \phi/2$  is gate tunable. Further, the scaling of  $\phi$  and the device conductance  $G_0$  also provides experimental knobs with which to adjust the responsivity of the device.

The optimal responsivity occurring at  $k_B T_g \approx \phi/2$  is an important characteristic for the design of graphene photodetectors. Indeed, given that  $T_g < 2000\text{ K}$  in realistic situations, Schottky barriers in the 100 meV ballpark allow operation of the g/X photodetector near optimal responsivities (i.e. near minimum  $\zeta$ ). These values of  $\phi$  can be achieved, for instance, in graphene- $\text{WS}_2$  devices [10].

Although g/X photodetectors allow in situ control of  $\phi$  by electrostatic doping, it is important to note that several parameters of the model vary implicitly with  $\phi$ . On the one hand, changes in  $\phi$  also induce changes in graphene doping, thus modifying the electronic cooling power. Further, when  $\phi$  becomes smaller than the incoming photon energies, photo-emission of primary carriers over the barrier competes with thermalization by electron-electron interactions. In this case, a smaller amount of the incident power is captured in the hot-carrier distribution.

Naturally, there are other mechanisms for losses that affect the responsivity. For instance, lateral (in-plane) heat currents,  $\mathcal{J}_q^\parallel = -\nabla \cdot (\kappa_\parallel \nabla T_g)$ , can transport heat towards the contacts in small devices. To estimate this

effect, we use the Wiedemann-Franz relation,  $\kappa_\parallel(T_g) = (\pi^2/3e^2) \times k_B^2 T_g \sigma$ , where  $\sigma$  is the in-plane electrical conductivity of graphene. For the relevant regime of moderate to high temperatures,  $T_g \gtrsim \phi$ , we can approximate  $\mathcal{J}_q^\perp \approx G_0 k_B^2 T_g^2 / e^2 = \gamma^\perp(T_g) \times T_g$  [cf. Eq.(2)]. As a result, we find a cooling length  $\xi_\parallel = \sqrt{\kappa(T_g)/\gamma^\perp(T_g)}$  coming from the thermionic channel that is independent of  $T_g$ . Using a uniform in-plane  $\sigma \sim 1\text{ mS}$  [27], we find  $\xi_\parallel \approx 0.6\text{ }\mu\text{m}$ , so that vertical energy extraction dominates over in-plane thermal conduction for sufficiently large devices with size  $L > \xi_\parallel$ .

We note that interactions with the substrate can result in cooling via surface optical phonons. These losses will vary for different substrate (X) choices and are only significant when X is a polar material [28]. Importantly, we do not expect them to be relevant in non-polar materials, e.g. X = silicon.

Lastly, it is interesting to note that g/X photodetectors can also operate at low photon energies,  $\hbar\omega \leq 2\mu$ . In this regime, conventional Drude absorption from ambient carriers directly captures incident radiation. This contrasts with conventional semiconductor photodetectors, that do not absorb light below the semiconductor bandgap. A tantalizing possibility is to use g/X Schottky junctions within the mid IR - THz bandwidth where presently-available technologies offer lackluster performance [29, 30].

In summary, graphene Schottky junctions host tunable interfaces across which energy transport can be engineered, exemplified by thermionic-dominated transport regime wherein energy and charge currents are strongly coupled. Fingerprints of the thermionic-dominated regime include high responsivities on the order of  $\mathcal{R} \sim 1 - 10\text{ A/W}$ , and a non-monotonic dependence of  $\mathcal{R}$  on electron temperature (or pump power) in g/X photodetectors. The large degree of in situ tunability allows optimization of the g/X interface for different applications and irradiation conditions; vertical hot carrier convection opens up new vistas to efficiently harvest photon energies over a wide spectral range, utilizing the entire exposed graphene area as a photoactive region.

We are grateful to useful discussions with M. Baldo, M. Kats, L. Levitov. We also thank V. Fatemi, A. Frenzel, and K. Tielrooij for a critical reading. JFRN and MSD acknowledge financial support from the National Science Foundation Grant DMR-1004147. JCWS acknowledges support from a Burke Fellowship at Caltech.

*JCWS Current Address:* Institute of High Performance Computing Singapore, and Division of Physics and Applied Physics, Nanyang Technological University.

- [2] X. Miao, S. Tongay, M. K. Petterson, K. Berke, A. G. Rinzler, B. R. Appleton, A. F. Hebard, *Nano Letters* **12**, 2745 (2012).
- [3] X. Wang, Z. Cheng, K. Xu, H. K. Tsang, J.-B. Xu, *Nat. Photonics* **7**, 888 (2013).
- [4] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, M. I. Katsnelson, L. Eaves, S. V. Morozov, A. S. Mayorov, N. M. R. Peres, A. H. C. Neto, J. Leist, A. K. Geim, L. A. Ponomarenko, K. S. Novoselov, *Nano Letters* **12**, 1707 (2012).
- [5] W. J. Yu, Z. Li, H. Zhou, Y. Chen, Y. Wang, Y. Huang, Xiangfeng Duan, *Nat. Materials* **12**, 246 (2013).
- [6] X. Li, H. Zhu, K. Wang, A. Cao, J. Wei, C. Li, Y. Jia, Z. Li, D. Wu, *Adv. Mater.* **22**, 2743 (2010).
- [7] C.-C. Chen, M. Aykol, C.-C. Chang, A. F. J. Levi, S. B. Cronin, *Nano Letters* **11**, 1863 (2011).
- [8] H. Yang, J. Heo, S. Park, H. J. Song, D. H. Seo, K.-E. Byun, P. Kim, I. Yoo, H.-J. Chung, K. Kim, *Science* **336**, 1140 (2012).
- [9] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, N. M. R. Peres, J. Leist, A. K. Geim, K. S. Novoselov, L. A. Ponomarenko, *Science* **335**, 947 (2012).
- [10] T. Georgiou, R. Jalil, B. D. Belle, L. Britnell, R. V. Gorbachev, S. V. Morozov, Y.-J. Kim, A. Gholinia, S. J. Haigh, O. Makarovskiy, L. Eaves, L. A. Ponomarenko, A. K. Geim, K. S. Novoselov, A. Mishchenko, *Nat Nano* **8**, 100 (2013).
- [11] A. Di Bartolomeo, *Physics Reports* **606**, 1 (2016).
- [12] R. Bistritzer, A. H. MacDonald, *Phys. Rev. Lett.* **102** 206410 (2009).
- [13] W.-K. Tse, S. Das Sarma, *Phys. Rev. B* **79**, 235406 (2009).
- [14] J. C. W. Song, M. Y. Reizer, L. S. Levitov, *Phys. Rev. Lett.* **109** 106602 (2012).
- [15] M. W. Graham, S.-F. Shi, D. C. Ralph, J. Park, P. L. McEuen, *Nature Physics* **9**, 103 (2013).
- [16] A. C. Betz, S. H. Jhang, E. Pallecchi, R. Ferreira, G. Feve, J.-M. Berroir, B. Placais, *Nature Physics* **9**, 109 (2013).
- [17] N. M. Gabor, J. C. W. Song, Q. Ma, N. L. Nair, T. Taychatanapat, K. Watanabe, T. Taniguchi, L. S. Levitov, P. Jarillo-Herrero, *Science* **334**, 648 (2011).
- [18] Koppens, F. H. L., Mueller, T., Avouris, Ph., Ferrari, A. C., Vitiello, M. S., Polini, M., *Nat Nano* **9**, 780 (2014).
- [19] K. J. Tielrooij, J. C. W. Song, S. A. Jensen, A. Centeno, A. Pesquera, A. Zurutuza Elorza, M. Bonn, L. S. Levitov, F. H. L. Koppens, *Nature Physics* **9**, 248 (2013).
- [20] J. C. W. Song, K. J. Tielrooij, F. H. L. Koppens, L. S. Levitov, *Phys. Rev. B* **87** 155429 (2013).
- [21] S. M. Sze, *The physics of semiconductor devices*, Wiley, New York (2007).
- [22] R. R. Nair, P. Blake, A. N. Grigorenko, K. S. Novoselov, T. J. Booth, T. Stauber, N. M. R. Peres, and A. K. Geim, *Science* **320**, 1308 (2008).
- [23] K. Rajkanan, R. Singh and J. Shewchun, *Solid-State Electronics* **22**, 793 (1979).
- [24] H. Yan, T. Low, W. Zhu, Y. Wu, M. Freitag, X. Li, F. Guinea, P. Avouris and F. Xia, *Nat Photon* **7**, 394 (2013).
- [25] J. F. Rodriguez-Nieva, M. S. Dresselhaus, and L. S. Levitov, *Nano Letters* **15**, 1451 (2015).
- [26] J. R. Goldman, and J. A. Prybyla, *Phys. Rev. Lett.* **72** 1364 (1994).
- [27] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos and A. A. Firsov, *Nature* **438**, 197 (2005).
- [28] A. S. Price, S. M. Hornett, A. V. Shytov, E. Hendry, D. W. Horsell, *Phys. Rev. B* **85**, 161411 (2012).
- [29] W. L. Chan, J. Deibel, and D. M. Mittleman, *Reports on the Progress of Physics* **70** 1325 (2007).
- [30] A. Rogalski, *Infrared Detectors*, CRC Press: Boca Raton, Florida (2007).